DOCUMENT RESUME

ED 276 399 IR 012 387

AUTHOR Woehlke, Paula Lee
TITLE An Evaluation of SYSTAT 2.1.
PUB DATE 19 Apr 86
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Software; *Microcomputers; Professional Occupations; *Statistical Analysis

ABSTRACT

This report provides a detailed description and evaluation of the SYSTAT 2.1 microcomputer statistics system for professional statisticians. Information provided includes the producer, versions available and compatibility, and a description of each of the 11 modules that make up the SYSTAT package. Discussion of the usability of the software includes descriptions of SYSTAT's 417-page documentation manual and three manual supplements, information on the ease of using and running programs, and information on program output, accuracy, and user services. It is concluded that SYSTAT 2.1 is fast, provides clear output, has routines for completing most common analytic procedures, is reasonably priced, and has adequate support available. While it is not recommended as a teaching tool, it is suggested that the system could be used in conjunction with a statistics course as a good example of the power of microcomputers for statistical analysis. It is recommended for those looking for a general purpose statistical package for the most common brands of microcomputers. Nine figures include examples of SYSTAT's numerous help screens and output, and graphic output from a standard IBM graphics printer. (DJR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED276399

An Evaluation of SYSTAT 2.1

Paula Lee Woehlke, Ph. D. Southern Illinois University-Carbondale

Paper Presented at the Annual Meeting of the
American Educational Research Association, San Francisco, CA
April 19, 1986

IR012387

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Paula Lee Woehlke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

An Evaluation of
SYSTAT 2.1

General Information

SYSTAT is available from SYSTAT, Inc., 603 Main Street, Evanston, IL 60202 (Phone 312/864-5670). Versions are available for the IBM-PC and compatibles using MS-DOS 2.0 or higher and with at least 256K of memory and double-drives or a hard disk ($495), CP/M machines with 64K of memory and 128K of disk storage ($495; Apple, HP 150, 8" format, and DEC versions cost an additional $25), and the Macintosh with 512K memory ($495; an additional disk drive is recommended). Versions are also available for other machines including Alpha Micro and HP 9000, with prices given upon request. The IBM-PC version of the program is contained on five diskettes; the DATA module is copy protected, although an unprotected version can be obtained for $10 and the return of a signed Single User License Agreement. The remaining disks (CLUSTER, NPAR, SERIES; CORR, MGLH; GRAPH, STATS, TABLES; and FACTOR, MDS) are unprotected. Compatible modules available at extra cost include a large (512K) version for IBM-PCs and compatibles ($50), Report Writer ($60), Non-linear Estimation ($100 at present, but to be included in SYSTAT 3 in the spring of 1986), PROBIT ($85), and STAT/TRANSFER, which permits one to download data from a mainframe or use data from LOTUS 1-2-3 ($65). A shipping charge of $5 ($20 for next day air or $10 for second day air) is also imposed on each order.

SYSTAT was developed by Leland Wilkinson; Mark Bjerknes developed the full-screen data editor available in the MS-DOS version. SYSTAT was first introduced in 1984.

The SYSTAT package consists of eleven modules. The DATA module includes routines for entering, merging and editing data (full screen version available on IBM-PCs and compatibles only). GRAPH permits the user to create histograms, box-and-whisker plots, scatterplots, function plots, stem-and-leaf plots, and normal probability and quantile plots of data. STAT obtains basic statistics (counts, minima, maxima, means, standard deviations, skewness, kurtosis, standard error of the mean, range, sum), and Tukey, Duncan, and Newman-Keuls post hoc tests. TABLES creates contingency tables and calculates associated statistics. CORR computes Pearson, Gamma, Guttman mu2, Spearman, and Kendall tau-b coefficients and sums-of-squares, covariance, and Euclidean distance matrices. MGLH calculates simple, multiple, logistic, and stepwise regression, univariate and multivariate analysis of variance, univariate and multivariate repeated measures analyses of variance, analysis of covariance, polynomial regression and trend surface analysis, principal components analysis, canonical correlation analysis, and partial correlations. FACTOR performs principal components analysis and provides plots of factor loadings, rotations, and factor scores. MDS does multidimensional scaling using either Kruskal's or Guttman's approach. CLUSTER performs cluster analyses with

options available for distance metrics, clustering method, and type of graphic display. NPAR computes such nonparametric statistics as sign, Wilcoxon, Friedman, Kruskal-Wallis, and Kolmogorov-Smirnov tests. Finally, SERIES performs time-series analyses.

Software Usability

Documentation

Documentation for SYSTAT consists of a 417-page manual and three manual supplements: "Read Me First," containing instructions on preparing the program to run on the particular computer system you have, a manual supplement of information on version 2.1 of the program, and a reference card. The manual is in an 8" x 9" looseleaf binder, with two plastic pages to hold diskettes in the front and each section separated by a tabbed heavy paper divider. The appendices and index are also separated from the other sections by a tabbed divider. The appendices include a command summary, elementary examples of the procedures, documentation of SYSTAT files (which are written in FORTRAN), and instructions for customizing the full screen editor. At the front of the manual, an introduction and tutorial take one through the use of SYSTAT to make graphs, descriptive statistics, and simple regression on a large data set of U.S. crime statistics included on the DATA module. The next eleven sections present the features of each of the SYSTAT modules. The DATA module is the most extensive section, consisting of four chapters. The first of these explains the use of the full screen data editor, available on the IBM-PC version of the package; the second illustrates the collection of data from various sources, such as the keyboard and ASCII files; the third describes the use of SYSTAT BASIC, a programming language built into SYSTAT to enable one to recode missing values, create value labels, create grouping variables, recode values, find outliers, generate random variables, etc.; and the fourth chapter is a general reference and in addition contains advanced topics pertaining to the DATA module, such as the INPUT, RUN, SORT, STANDARDIZE, TRANSPOSE, and USE commands. The sections which follow document the use of the each module, and provide examples of each procedure and descriptions of the use of each command. For example, section three describes the options available in the STATS module, with examples; the only problem I found with this section was that one had to infer the existence of some optional output from the examples, as they were never explicitly described here (e.g., SUM as an option is mentioned in passing in an example; otherwise, one must refer to Appendix I to find it). Each section similarly describes and illustrates the use of some particular module of SYSTAT, generally clearly and in an easily read fashion. However, nowhere in the manual or promotional documentation does one find

descriptions of the algorithms used; comparisons of SYSTAT's accuracy with results of some mainframe packages is available in promotional materials and calls to the company would probably produce some ideas of the computing algorithms employed, although I have yet to try this.

Ease of Use

SYSTAT is moderately easy to use. Because of the systems on which it runs, there are system commands to be learned (e.g., responding "B:DATA" when confronted by the CP/M and MS-DOS "A>" prompt when one wants to load the data module) as well as the SYSTAT commands. A certain amount of disk-swapping is necessary, because only one to three modules will fit on a diskette. For example, in the IBM-PC version there are five diskettes to be swapped in and out of the B-drive while the user's "work" or data diskette resides in the A-drive.

Tutorial. SYSTAT contains no tutorial diskette; however, it does come with a sample data set for use in the tutorial section of the manual. Because of the necessary level of statistical expertise of the intended user of the program, a tutorial diskette would not seem to be particularly useful.

Help Screen. SYSTAT includes numerous help screens throughout the modules, examples of which can be found in Figures 1 to 3. There is an overall help screen for each module as well as screens pertaining to subcommands within the modules. They generally describe syntax of commands and do not aid the statistically naive in determining the correct procedure or understanding output.

Menu Driven or Command Driven. SYSTAT is command-driven, which is both a blessing and a curse. It is a blessing in that one can bypass unnecessary information in creating a statistical analysis; it is a curse because each module of the program can use from eleven to thirty commands to describe an analysis. Although some of these commands are common to all modules (e.g., HELP, SUBMIT, OUTPUT, PAGE, BY, QUIT), without the Help Screens mentioned above it would be extremely difficult to use the program because of the number and variety of commands. Formats of commands also differ among modules, another reason that the Help Screens are needed. However, syntax of most commands is not unexpected; it follows closely some of the more popular mainframe computer packages, and is easy to learn, given moderate usage of the package.

Error Diagnostics. Error messages when running SYSTAT are generally informative, but not always. For example, an attempt to issue a "HELP" command from within the EDIT submodule resulted in the message, "ERROR: YOU CAN ENTER COMMANDS ONLY ON LINE BELOW WINDOW" (it is assumed that you know how to get out of the window). However, issuing the command "B:FACTOR" to access the Factor module before two "quit" statements had been issued from within the Data module (one to quit the Edit

submodule and one to quit the Data module) resulted in the message, "ERROR: UNABLE TO DECIPHER TEXT AT END OF THIS LINE: B," which is considerably more cryptic than the former message. In such cases, experimentation is necessary to determine what went wrong. The manual itself has no summary of error messages; a short section describes only those errors which result in the user being bumped back to the operating system (e.g., disk full, input record too long), but calls all other error messages "self-explanatory."

Diskette Swapping. As noted earlier, a certain amount of disk-swapping is necessary to run the program unless one has a hard disk attached to the system. In general, this is not too annoying because a number of modules and submodules reside on each of the five diskettes. The only irritating aspect of the disk-swapping process is that when one leaves one module of SYSTAT to go to another module, the program re-initializes itself and the user must watch the SYSTAT logo appear on the screen, re-read the copyright information, and re-access the data set in use on the work disk.

Running Programs

Overall, SYSTAT deserves a high rating with respect to data manipulation and transformation, tests of assumptions, and currentness and appropriateness of algorithms.

Data Entry. For the IBM-PC version of SYSTAT, data are entered much as one would put data into a spreadsheet program. The screen takes on the appearance of a spreadsheet, and cursor movements and data entry procedures resemble those of, say, Visicalc or Supercalc. Thus data entry is an easy task for most users, as are correction of errors and inclusion of additional data values.

Data Manipulation and Transformation. Numerous commands are available for manipulating and transforming data from within the Data module. These include the abilities to append files, drop variables, get data from or write data to an ASCII file, retain values from one case to another within a file, perform both conditional and unconditional transformations (with IF and LET statements, respectively), rank, sort, and standardize data, transpose matrices, lag data, and perform multiple transformations. Standard transformation available through the "LET" statement include addition, subtraction, multiplication, division, exponentiation, comparison operators (=, <, >, etc.), logical operators (AND, OR), and single-variable transformations using square root, logarithm, integer truncation, normal, inverse normal, trigonometric, and inverse trigonometric functions.

Currentness of Algorithms. It is difficult to judge the currentness of algorithms from information provided in the documentation. However, I have found the results of SYSTAT runs to match output from SAS and SPSSX, and because the program is

written in FORTRAN using algorithms from mainframe packages, it can be assumed that only currently favored algorithms have been used.

Program Output

Output from SYSTAT is one of its best features. The output from a given command is printed immediately after the command is executed; in this way, the user can order the logical sequence of steps for the analysis. The format of similar output from different modules is generally consistent, although the contents of the modules is such that very few items of output appear in more than one of them. Every piece of output information is clearly labeled, and it is assumed that the user knows the meaning of the label (i.e., no explanations accompany the results). It is possible to print the results of analyses in either 80 or 132 columns on standard 8.5" by 11" paper through keyboard control of the printer. A facility is also provided for making comments on the printed output, as is a means whereby one may send output to a disk instead of the screen or printer. One can control the format of outputted numbers and output page width through the program, too. Examples of SYSTAT output are provided in Figures 4 and 5. Graphic displays are presented in easily readable format, using the highest level of graphic capability of the screen or printer (see Figures 6 through 9 for output from a standard IBM graphics printer [Epson RX series]).

Accuracy

As noted earlier, results of several SYSTAT analyses were compared with those obtained from SPSSX and SAS; results were the same. This is not to suggest that all results will be identical, because it was not possible to test every module for this review. However, according to the most recent SYSNET (Users' Newsletter), any bugs that are detected can be reported to SYSTAT, Inc., on a technical support telephone line available from 9 a.m. to 5 p.m. Central time, and staff will try to determine the cause, find a cure, or suggest a way to bypass the problem.

User Services

As noted above, SYSTAT provides a technical support line (not toll-free) 40 hours per week for registered owners of the package. In addition, users receive the aforementioned SYSNET newsletter, which contains reports of and corrections for bugs, information about update options, and techniques to make use of the program easier. Although no training sessions are advertised in either promotional materials or manuals, the company is encouraging the formation of users' groups. In addition, users can talk to SYSTAT representatives in the exhibit areas at most professional meetings.

Summary

SYSTAT appears to be a flexible microcomputer statistics system for professional statisticians. It is fast, provides clear output, and has routines for completing most common analytic procedures. In addition, the cost is reasonable for such a powerful package, and availability of support seems to be adequate. I would not recommend it as a teaching tool, but it could be used in a statistics course as a good example of the power of the microcomputer for statistical analysis; for most courses, it could be used in lieu of a mainframe package. I recommend it for those looking for a general purpose statistical package for the most common brands of microcomputers.

```
>help
     The DATA module processes and creates SYSTAT files.
     DATA commands are:
        EDIT
        INPUT
        TYPE, DIAGONAL
        RSEED, REPEAT
        USE, SAVE
        GET, PUT
        RUN
        HOLD, APPEND
        IF, LET
        RANK, SORT, STANDARDIZE, TRANSPOSE
        SUBMIT, OUTPUT
        NOTE, PAGE, FORMAT, HELP
        NEW, BY
        ERASE, DROP, LIST, LRECL
        ARRAY, CODE, LABEL
        DOS, QUIT
     For  additional  information about  these  commands,  type
     HELP RUN, etc.
>
help
     The DATA module processes and creates SYSTAT files.
     DATA commands are:
        EDIT
        INPUT
        TYPE, DIAGONAL
        RSEED, REPEAT
        USE, SAVE
        GET, PUT
        RUN
        HOLD, APPEND
        IF, LET
        RANK, SORT, STANDARDIZE, TRANSPOSE
        SUBMIT, OUTPUT
        NOTE, PAGE, FORMAT, HELP
        NEW, BY
        ERASE, DROP, LIST, LRECL
        ARRAY, CODE, LABEL
        DOS, QUIT
     For  additional  information about  these  commands,  type
     HELP RUN, etc.
>help edit
     The EDIT command starts the SYSTAT full screen editor.

     Its syntax is:

     EDIT <filename>

     Examples are:

     EDIT           (edit a new file)
     EDIT CRIME     (edit CRIME.SYS)
```

Figure 1. SYSTAT HELP screens from the DATA module

```
>help
     MGLH estimates and tests univariate and multivariate
     general linear models using data in a SYSTAT file.

     MGLH commands are:

        CATEGORY
        MODEL
        STEP
        ESTIMATE
        HYPOTHESIS
        PRINT
        USE, SAVE
        SUBMIT, OUTPUT
        NOTE
        PAGE
        FORMAT
        HELP
        BY, SELECT
        DOS, QUIT

     For additional information about these commands, type
     HELP followed by the name of the command.
>help hypothesis
     The HYPOTHESIS command begins the specification of a hypo-
     thesis on an estimated model.  A number of commands are
     used to construct a hypothesis.  Help with these commands
     is available after you estimate a model, type HYPOTHESIS,
     and then type HELP.
>e help esto,ate     imate
     The ESTIMATE command causes the previously specified model
     to be estimated.

     Just type:

     ESTIMATE
```

Figure 2. SYSTAT HELP screens from the MGLH module

```
>help
     The FACTOR module performs a principal components analysis
     of a COVARIANCE or CORRELATION matrix in a SYSTAT file and
     optional orthogonal rotation and factor scores.

     FACTOR commands are:

        TYPE
        SORT
        EIGEN, NUMBER, ROTATE
        FACTOR
        USE
        SAVE
        SUBMIT
        OUTPUT
        NOTE
        PAGE
        FORMAT
        HELP
        BY, SELECT
        DOS, QUIT

     For  additional  information about  these  commands,  type
     HELP followed by the name of the command.
>help factor
     The FACTOR command initiates factoring.
     Its syntax is:

     FACTOR [<var1>,<var2>,<...> / PLOT]

     Examples are:

     FACTOR  (all numerical variables in file are factored)
     FACTOR SIMILAR,BLOCK,PICTURE,VOCAB,ARITH
     FACTOR MMPI(1-10) / PLOT
```

Figure 3. SYSTAT HELP screens from the FACTOR module

```
>model rape=constant+assau;lt   lt+murder+robbery
>estimate

DEP VAR:     RAPE       N:    50     MULTIPLE R:  .723     SQUARED MULTIPLE R:  .523

                2          2
ADJUSTED R  = 1-(1-R )*(N-1)/DF, WHERE N=  50, AND DF=  46:  .491

  VARIABLE    COEFFICIENT    STD. ERROR    STD. COEF.  TOLERANCE      T     P(2 TAIL)

CONSTANT           5.538         1.676         0.000   .           3.30        .002
 ASSAULT           0.066         0.022         0.617   0.24714     3.01        .004
  MURDER          -0.110         0.344        -0.057   0.31985     -.32        .751
 ROBBERY           0.018         0.010         0.224   0.64280     1.76        .085

                                ANALYSIS OF VARIANCE

   SOURCE    SUM-OF-SQUARES    DF   MEAN-SQUARE      F-RATIO      P

 REGRESSION       1382.782      3       460.927        16.787    .000
 RESIDUAL         1263.025     46        27.457

>hypothesis
>effect=murder&robbery
>test

HYPOTHESIS FOR EFFECT CALLED:
                    MURDER
               AND
                    ROBBERY


A MATRIX

                          1           2           3           4

              1       0.000       0.000       1.000       0.000
              2       0.000       0.000       0.000       1.000


HYPOTHESIS SUM OF SQUARES

                      101.361


ERROR SUM OF SQUARES

                     1263.025

UNIVARIATE F TESTS
  VARIABLE          F(  2,   46)       PROBABILITY

      RAPE              1.846              .169
```

Figure 4. Output from SYSTAT MLGH multiple regression analysis

```
>model rape=constant+assault+murder+robbery
>step

STEPWISE REGRESSION WITH ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150

STEP=  1     ENTER  ASSAULT     R=  .696     RSQUARE=  .484
STEP=  2     ENTER  ROBBERY     R=  .722     RSQUARE=  .522

THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:

        CONSTANT
         ASSAULT
         ROBBERY

USE THESE PREDICTORS IN A NEW MODEL SENTENCE TO ESTIMATE THE COEFFICIENTS.

>model rape=assult      constant+assault+robbery
>esto,a   imate

DEP VAR:     RAPE       N:    50     MULTIPLE R:  .722     SQUARED MULTIPLE R:  .522

          2        2
ADJUSTED R = 1-(1-R )*(N-1)/DF, WHERE N=  50, AND DF=  47:  .501

  VARIABLE    COEFFICIENT    STD. ERROR    STD. COEF.  TOLERANCE      T     P(2 TAIL)

CONSTANT            5.467         1.646        0.000          .       3.32       .002
 ASSAULT            0.061         0.013        0.564    0.68280       4.62       .000
 ROBBERY            0.019         0.010        0.234    0.68280       1.91       .062

                              ANALYSIS OF VARIANCE

   SOURCE    SUM-OF-SQUARES    DF   MEAN-SQUARE       F-RATIO      P

 REGRESSION       1379.985      2       689.993         25.619    .000
 RESIDUAL         1265.822     47        26.932

-------------------------------------------------------------------------------

>category region=8
>model rape=region
>estimate

MODEL CONTAINS NO CONSTANT.

DEP VAR:     RAPE       N:    50     MULTIPLE R:  .374     SQUARED MULTIPLE R:  .140

                              ANALYSIS OF VARIANCE

   SOURCE    SUM-OF-SQUARES    DF   MEAN-SQUARE       F-RATIO      P

   REGION         2075.894     7       296.556            .999    .445

 ERROR           12762.886    43       296.811
```

Figure 5. Output from SYSTAT MLGH stepwise regression analysis and analysis using a categorical predictor variable

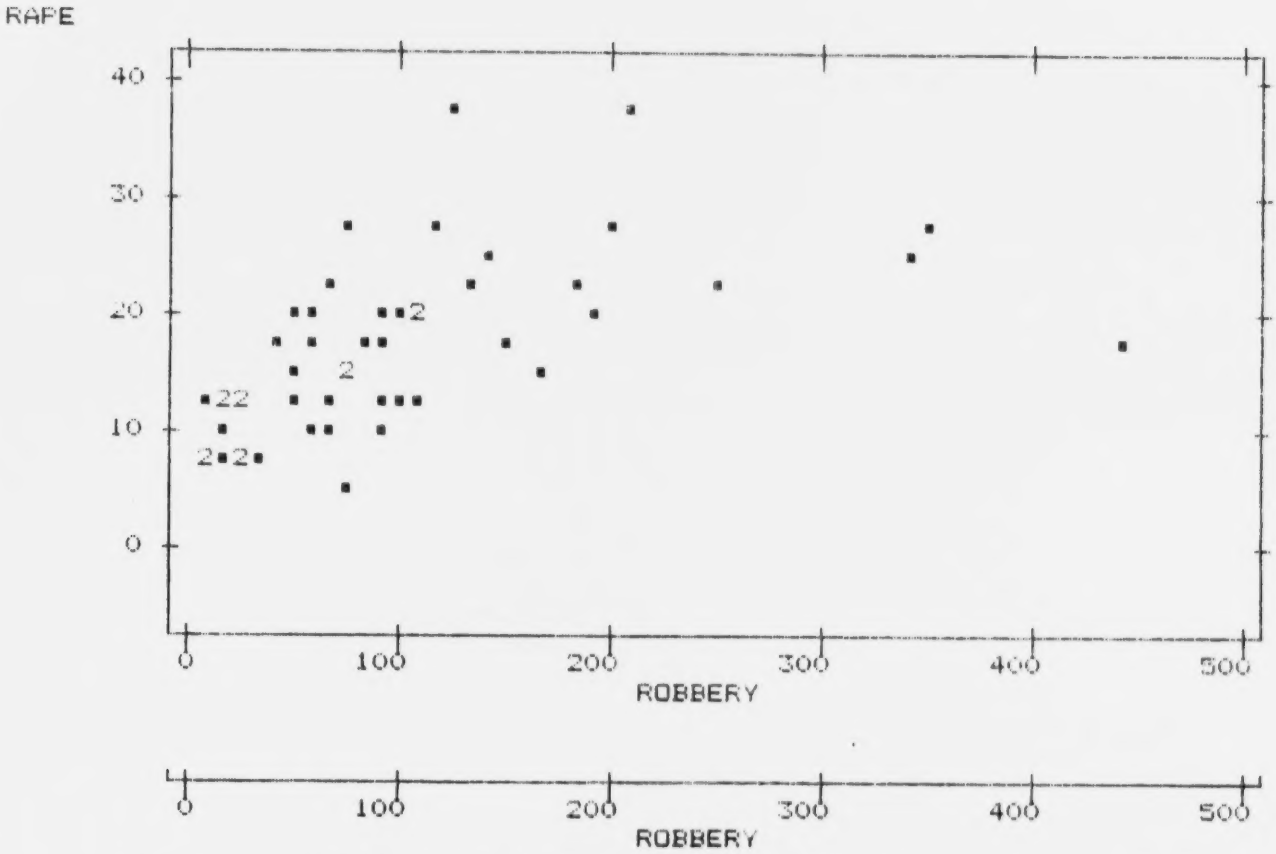


Figure 6. SYSTAT histogram and scatter diagram

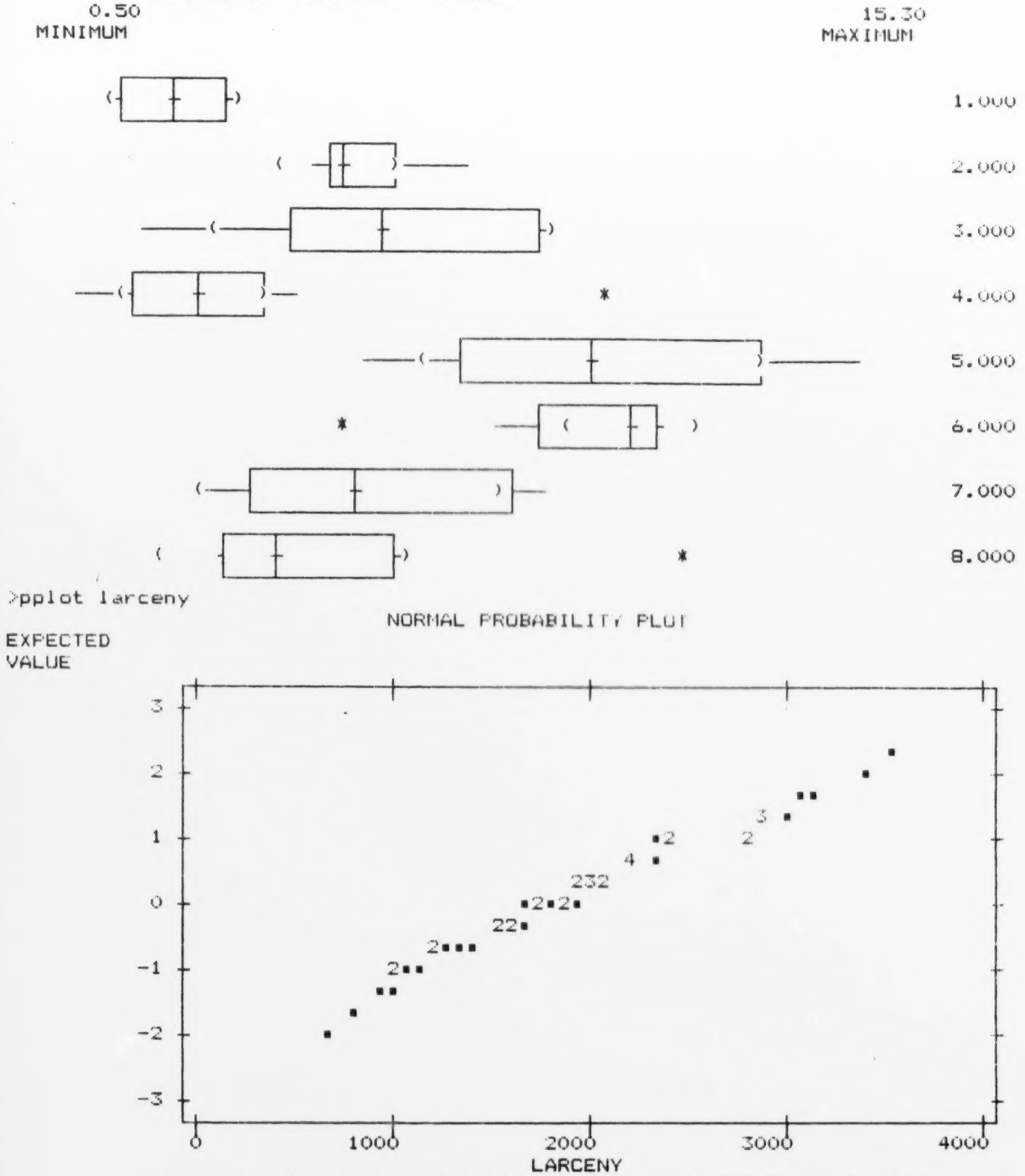
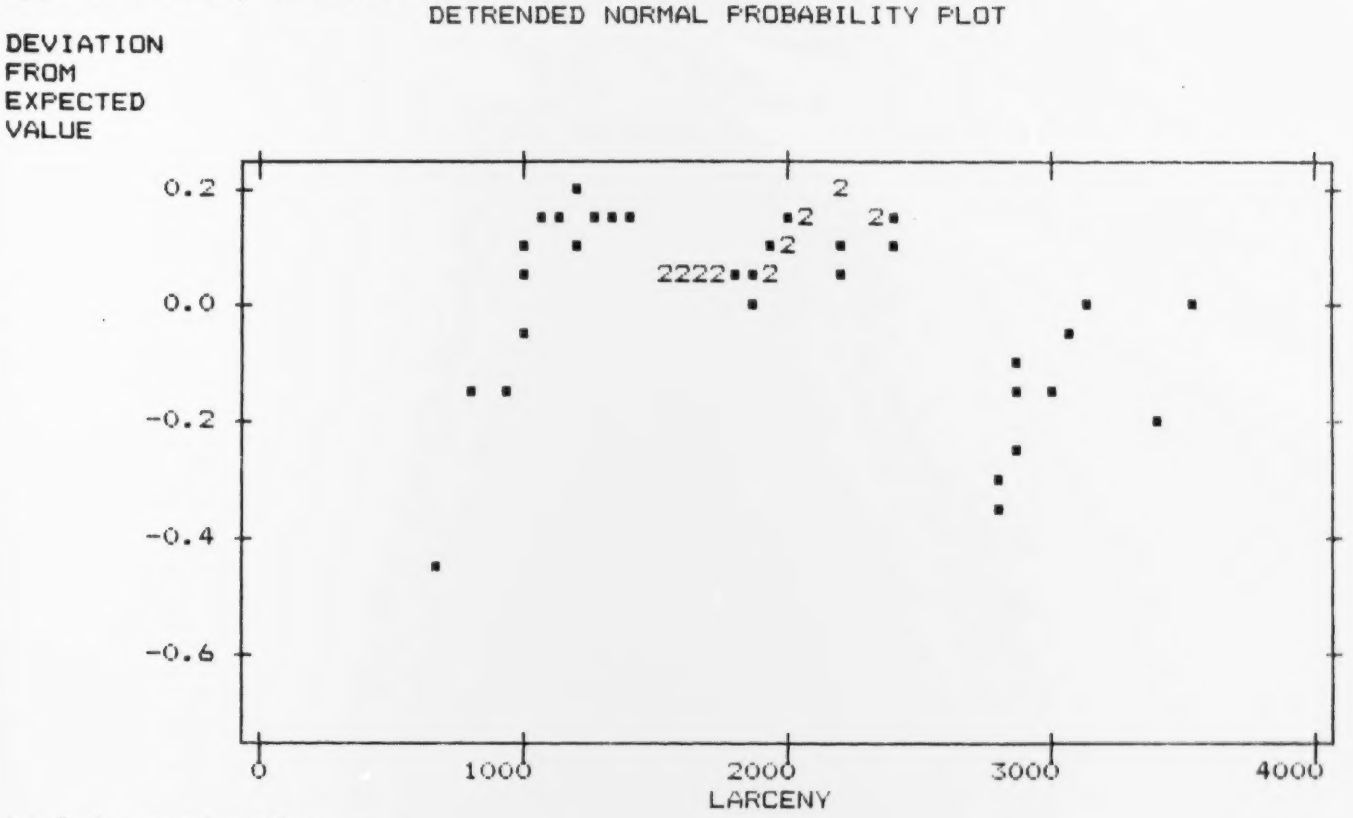


Figure 7. SYSTAT box-and-whisker and normal probability plots

```
>pplot larceny/detrend
```




```
>qplot rape*murder
```

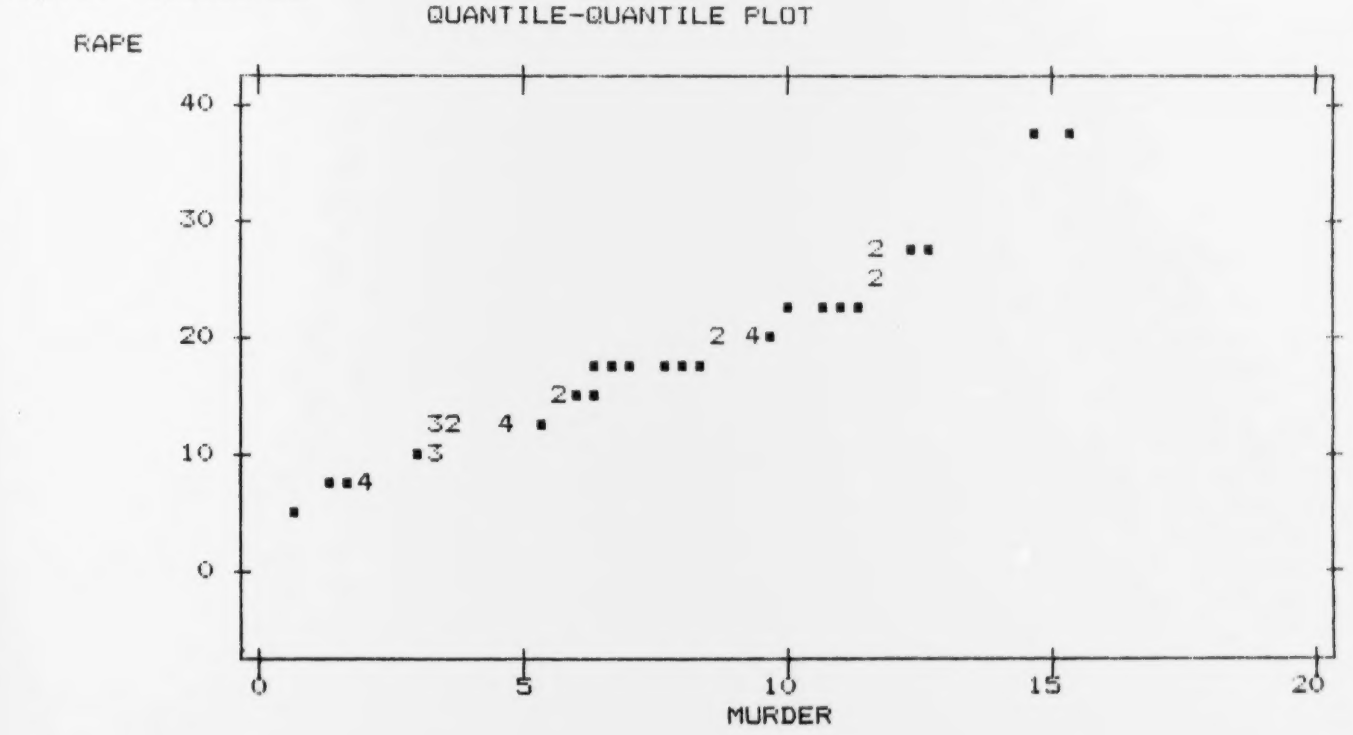


Figure 8. SYSTAT detrended normal probability and quantile-quantile plots